



openRGD – steg 2

Artikeln om openRGD i förra numret handlade om validering och kontroll av en Gedcom-fil. Nu tar vi upp möjligheten att använda ytterligare funktioner i openRGD för att bearbeta två olika Gedcom-filer.

NÄR MAN samarbetar med någon annan släktforskare finns det tillfällen när man skulle vilja jämföra släktdata. Att hitta gemensamma personer och gemensamma släktgrenar kan vara knepigt om man inte har turen att ha ovanliga släktnamn gemensamt.

I openRGD:s matchningsfunktion jämförs personerna i två Gedcom-filer på ett avancerat sätt och resultatet visas i kompletta familjebilder. Med hjälp av färgsättning och olika status ger det en direkt indikering på vilka personer och familjer man har gemensamt. Dessutom får man en tydlig indikering på olikheter i familjerna.

I vilka lägen har man då intresse av att jämföra två släktforskningar med varandra? Det vanligaste är troligen att man av en slump hittat en annan släktforskare på något forum, i Disbytt eller någon annanstans på nätet. Man har hittat en gemensam ana och vill undersöka om det finns fler kopplingar.

Om man har gjort en DNA-analys får man plötsligt massor av släktingar, men det är inte alltid så lätt att komma på hur man är släkt. Men om båda parter har var sin släktdata kan openRGD vara en bra genväg för att hitta kopplingen.

Ett annat exempel, som många gånger ställer till stora problem, är att man tar in uppgifter man hittat på nätet eller via en Gedcom-fil till sin släktdata. Att plocka in den typen av uppgifter direkt i sin släktforskning är något man aldrig bör göra. Man kan få in uppgifter med dålig kvalitet men det allvarligaste är att det kan skapa dubletter som kan bli svåra att få bort.

Här är openRGD ett utmärkt hjälpmedel, dels får man en kvalitetsbedömning av uppgifterna, dels får man en bra bild av hur väl de kompletterar ens uppgifter och hur stor mängd dubletter man riskerar att få.

De tre funktioner som openRGD erbjuder är:

1. Maskinell matchning, man talar bara om vilka som ska matchas och startar funktionen.
2. Manuell matchning, det är här man ser resultatet och det är här man gör en manuell bearbetning.
3. Visa skillnader, där kan man hitta skillnader i uppgifter för de matchade personerna.

Den manuella matchningen kan föranleda en del arbete om man vill fullfölja bearbetningen helt. Det går dock alldeles utmärkt att helt utan bearbetning få en bra snabböversikt genom att bara se resultatet från den maskinella matchningen.

I samband med att Gedcom-filen kontrollerades skapades också en databas som används vid matchningen. Men för att göra en matchning måste man ha tagit in minst två olika Gedcom-filer.

Om man loggat in som guest måste dessa göras direkt efter varandra för att båda ska vara tillgängliga för matchning. Om man bara vill göra en

4. Maskinell Matchning - [Läs mera](#)

Matching av två databaser innebär att programmet försöker att identifiera personer, som finns i båda databaserna.

Databas: att matchas mot:

5. Manuell Matchning - [Läs mera](#)

Den maskinella matchningen kan inte med tillräcklig säkerhet matcha alla personer mot varandra, utan linärar tveksamma matchningar till manuell bedömning.

Detaljerad dokumentation (visas i nytt fönster): [Matchinfo.pdf](#)

Databas: att matchas mot:
<möjliga databas II val>

6. Visa skillnader - [Läs mera](#)

Visar detaljskillnader mellan matchade personer.

RGD Familjelista: kalle_Test -> kalle_Kalle

[Byt till familj](#) || [Byt till person](#) || [Åtgärdslista](#) || [Tillbaka till startsida](#) || [Logga ut](#)

MultiMatch: DB1 ☐; DB2 ☐

Inkludera: Grupper allOK ☐; allManuell ☐; allEjOK ☐

Status: rOK ☐; OK ☐; Match ☐; rManuell ☐; Manuell ☐; EjOK ☐; rEjOK ☐; EjMatch ☐; FamEjOK ☐

|| Totalt: 0

Sök efter Gedcom ID: :

Bild 1. De tre valen för att jämföra Gedcom-filer på openRGD:s webbplats.

Bild 2. Startbild för den manuella matchningen.

snabbkontroll fungerar det bra, men vid en mer djupgående bearbetning är det fördelaktigt att ha ett eget konto och lösenord. Det är enkelt för Dis medlemmar att få, det är bara att höra av sig.

Maskinell matchning

Maskinell matchning är enkel, man väljer en av databaserna som databas I, den andra som databas II och sedan klickar man på knappen Matcha. Man skriver inte namnet, programmet talar snällt om vilka databaser man skapat och som kan väljas. Programmet talar också om när det är klart (loggen som visas har ingen funktion i detta läge). Snabbast resultat blir det om man som databas I använder den med minst antal personer. I länken "Läs mera" finns information om man är osäker.

Manuell matchning

Nästa steg är Manuell matchning där man anger samma databas I som tidi-

gare och sedan samma databas II. Har man angivit dessa korrekt dyker startknappen upp. Det är här i den manuella matchningen som informationen från den maskinella matchningen visas. På sidan finns även dokumentet "Matchinfo.pdf" som kan vara lämplig att läsa.

Startbilden innehåller många valmöjligheter, men ganska få behövs för att få en övergripande bild. Man arbetar med familjer eller med personer. När man startar är det familj som är aktiv.

Genom att markera rutan för "Inkludera: Grupper allOK", får man upp familjer som blivit matchade i den maskinella matchningen, se bild 3. Matchade personer är markerade med grön färg. Vid en snabbkontroll är detta det enda som behöver göras.

Manuell bearbetning

Den manuella matchningen går ut på att gå igenom alla de familjer/personer som den maskinella matchningen

hoppat över. I varje bild i den manuella matchningen finns ett antal knappar. En del ger tillgång till fler uppgifter och två knappar accepterar eller avslår en matchning. Första steget är alltid att eliminera eventuella MultiMatch. Multimatch betyder att en och samma familj eller person har fått mer än en träff vid matchningen. Det är därför mest naturligt att dessa kan bearbetas tillsammans för att få bäst översikt.

De många statusvärdena som finns i bilden kan användas i olika sammanhang för selektering men är inget man normalt behöver använda. Det finns också möjlighet till selektering om det rör sig om stora datamängder.

Om man upptäcker felaktigheter kan man i gruppen allOK bryta en maskinell matchning genom knappen "ej OK". I övrigt behöver man inte acceptera de som är riktiga.

Grupp "allManuell" innehåller personer där den maskinella matchningen inte haft tillräckligt underlag

Inkludera: Grupper allOK ☒; allManuell ☐; allEjOK ☐;
 Status: rOK ☒; OK ☒; Match ☒; rManuell ☐; Manuell ☐; EjOK ☐; rEjOK ☐; EjMatch ☐; FamEjOK ☐;

|| Totalt: 9

Sök efter Gedcom ID: i Databas I

#	Status	Ant	Id	Namn	Gift	
1	Match	1	gedcom_F2660	Anders Erik /Ersson/ Mathilda Sofia /Ersson f Andersson/	18830811 Källreferens till gruppen Sveriges dödbok 1901-2009 utan källa ovanför	Visa
2	Match	1	gedcom_F2711	Kerstin /Persdotter/		Visa
3	Match	1	gedcom_F2711	Kerstin /Persdotter/		Visa
4	Match	1	gedcom_F2721	Anders* Peter /Anderson/	1870 Åsker (T)	Visa

Bild 3. Svarsbilden för grupp "allOK" visar en rad för varje familj.

Familj

Roll	Status	Namn/relid	Född	Död	Namn/relid	Född	Död
husb	notMatched				Erie /Kosonen/ gedcom_26-13266	1773 *4 Madsjö AI.5, Sidas 466	
Flags					View		
wife	Match	Kerstin /Persdotter/ gedcom_B9099	1773 Brooks, Stora Malm (D) *4 Östra Vingåker (D) AI.5, Sidas 138	18390316 Nordlund, Östra Vingåker (D) *4 Östra Vingåker (D) AI.6, Sidas 223	Mina /Persdotter/ gedcom_26-11909	17731226 *4 Madsjö AI.4, Sidas 247	
Flags					View		
Marriage	Match	gedcom_F2711			gedcom_26-12374	18051113 Madsjö (H) *1 Madsjö CI.3, Sidas 591	
child					Sara Mina /Ericdotter/ gedcom_26-5038	18070908 Madsjö (H) *4 Madsjö AI.5, Sidas 466	
Flags					View		
child					Johannes /Ericsson/ gedcom_26-5053	18131115 Madsjö (H) *4 Madsjö AI.5, Sidas 466	
Flags					View		

Hela familjen OK || || Familjen _INTE_ OK

Bild 5. Felaktigt resultat av den maskinella matchningen.

för att avgöra om personerna matchar eller inte. De är då avsedda att bearbetas manuellt och är gulmarkerade. Man måste först gå igenom de olika familjebilderna och sedan de olika personerna. Man väljer då att med knapparna OK och ejOK markera match eller inte match. De markerade familjerna respektive personerna flyttas då till gruppen alltOK eller alltEjOK. Gruppen alltEjOK bearbetar man normalt inte alls.

När gruppen allManuell är tom både för familj och för person så är den manuella matchningen klar och det verkliga resultatet av jämförelsen definitivt fastställt.

Övriga statusar och funktioner används först när man gör mer avancerade bearbetningar.

Exempel manuell matchning

I svarsbilden för grupp "allOK" visas en rad för varje familj i den fil jag kallat Test, som hade motsvarande

match i min fil Kalle (se bild 3). I huvudet anges hur många familjer totalt som fått träff.

Genom att klicka på "Visa" syns båda familjerna under listningen (se bild 4). Bilden kan vid första anblicken se rörig ut men det är ganska enkelt att direkt se vilket resultat jämförelsen gett. Bilden består av två halvor, till vänster familjen från databas I och till höger familjen från databas II.

De två översta raderna i bilden är makarna, den tredje raden är information om relationen och därefter följande rader representerar barn.

Färgerna grön markerar match, för tveksamma kopplingar är raden gul och de mindre troliga är rödmarkerade. Unika personer i familjen saknar färgmarkering. Denna familj är en av de vi har gemensamt i vår forskning.

Man kan göra fler tolkningar av bilden. Man kan se att jag huvudsakligen har mer information som den andra släktforskaren kan dra nytta av,

till exempel ytterligare fyra barn. Om jag tagit in Gedcom-filen direkt i min forskning hade jag fått in tre dubletter. Alla personer har identiteter som gör att vi kan kommunicera med varandra.

Man måste också vara kritisk till resultatet, för även datorer kan göra fel. I vårt exempel har namn och årtal lurat datorn att koppla ihop två personer trots att de är födda i olika delar av landet (se bild 5).

Mina exempel visar hur man snabbt och utan större manuellt arbete kan få tillräcklig information för att avgöra om det är intressant att göra djupare analyser och jämförelser.

Visa skillnader

Ytterligare en bra funktion i openRGD är "Visa skillnader". Funktionen jämför matchade personer med varandra och visar om det finns skillnader. Det kan mycket väl finnas olika uppgifter, för matchningen är inte någon exakt

Familj

Roll	Status	Namn/refid	Född	Död		Namn/refid	Född	Död
husb	Match	Olof /Andersson/ gedcom_19399	1729 *4 Asker A:6, Sidan 101	17870421 Tovimult, Asker (T) *4 Asker A:6, Sidan 101	View	Olof /Andersson/ gedcom_26-12105	17290718 Asker (T) *1 Asker C:2, Sidan 15	17870421 Asker (T) *2 Asker F:1, Bild 110
wife	Match	Brita /Ersdotter/ gedcom_19400	1731	17870425 Tovimult, Asker (T) *2 Asker F:1, Bild 110	View	Brita /Ersdotter/ gedcom_26-13238	1731 Länås (T) *4 Asker A:5, Sidan 88	17870425 Asker (T) *2 Asker F:1, Bild 110
Marrings	Match	gedcom_F2825				gedcom_26-4473		
child					View	Anders /Olsson/ gedcom_26-2679	17580923 *4 Asker A:6, Sidan 101	
child					View	Erie /Olsson/ gedcom_26-13591	17600310 Asker (T) *1 Asker C:2, Sidan 159	18140117 Asker (T) *1 Asker F:2, Sidan 46
child	Match	Olof /Olsson/ gedcom_19397	17630701 Tovimult, Asker (T) *4 Asker A:6, Sidan 101		View	Olof /Olsson/ gedcom_26-396	17630901 *4 Asker A:5, Sidan 88	
child					View	Catharina /Olsdotter/ gedcom_26-13650	17670209 Asker (T) *1 Asker C:2, Sidan 225	17780603 Asker (T) *2 Asker F:1, Bild 90
child					View	Pär /Olsson/ gedcom_26-2815	17690302 *4 Asker A:6, Sidan 101	

Bild 4. Genom att klicka "Visa" i svarsbilden för grupp "allOK" visas båda familjerna från de olika filerna.

jämförelse. Personerna som har olika födelseplats eller datum bör kollas extra noga. Det kan både vara fel i matchningen eller i släktforskningen. Olika namn kan ge indikation på felstavningar.

Avser man att gå vidare med bearbetningen är den första åtgärden att gå igenom listorna från indatavalideringen. Att kontrollera och eventuellt rätta i sina data så långt det är möjligt och därefter skapa nya Gedcom-filer. Viktigt för kvaliteten på matchningen är att det inte finns dubletter och att alla personer har rätt kön. Även att församlingarna för född, död och vigd kunnat identifieras som församlingar eller län. Detta är inte lika viktigt men det påverkar också matchningskvaliteten.

Fler funktioner

Efter genomförd matchning går det att ta ut resultatet i en xlsx- eller en csv-fil. Det underlättar betydligt kommunikationen mellan de båda

släktforskarna. Vill man slå samman de båda Gedcom-filerna till en gemensam fil så finns det en funktion i openRGD även för detta. Vid en korrekt matchning görs sammanslagningen utan att dubletter skapas. En ny Gedcom-fil kan därefter skapas med alla unika data. För att fullfölja matchningen krävs dock manuellt arbete. Beroende på hur mycket som är gemensamt blir också mängden manuellt arbete mindre eller större.

Övning ger färdighet

Det är inte lätt att göra en beskrivning som täcker allt. Det bästa sättet är att börja jobba med två mindre Gedcom-filer som har några gemensamma uppgifter och praktiskt prova sig fram. Att steg för steg lära sig förstå de olika momenten.

När programmen skapades utfördes mängder av tester för att få ett logiskt arbetsflöde och mycket funktionalitet komprimerat i tydliga och lätthante-

rade bilder. Om det krävs mycket eller lite arbete beror på typen och mängden data, inte på programmets komplexitet.

Prova på och testa! Man behöver aldrig vara rädd att förstöra något, det går alltid att börja från början igen. När man har ett eget konto sköter man själv borttagning av databaser man inte längre använder. Långt ner på sidan finns en länk till de databaser man skapat. Där kan man ta bort de som inte längre används aktivt.□

MER INFORMATION

Det finns en länk till openRGD på Dis webbplats: www.dis.se/openrgd.